DOCUMENT RESUME

ED 126 127 TM 005 383

AUTHOR Anderson, Edwin R.
TITLE An Evaluation of the University of Washington Innovative Fund.
INSTITUTION Washington Univ., Seattle. Educational Assessment Center.
REPORT NO EAC-P-551
PUB DATE May 76
NOTE 27p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Administrative Policy; Curriculum Development; Educational Improvement; *Evaluation; *Grants; *Higher Education; *Instructional Innovation; Program Effectiveness; *Resource Allocations
IDENTIFIERS *University of Washington

ABSTRACT

The University of Washington committee on instructional development and innovation is evaluated with respect to their effectiveness as a granting agency. The paper begins with a section defining evaluation and then applies that definition in a conceptual and empirical critique of funded activity. Conclusions drawn in the report include: (1) evaluation should not be required of one-time funding for new course development, (2) the committee should provide the grant receivers with specific evaluation requirements, (3) the committee should accept the responsibility for evaluating the overall funding pattern they establish, (4) the conceptual base of media-related evaluation needs to be changed, and (5) funding for the improvement of existing courses should extend through more than one grant. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

An Evaluation of the University of Washington Innovative Fund

Edwin R. Anderson

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Educational Assessment Center

University of Washington

May 1976

An Evaluation of the University of Washington Innovative Fund

Edwin R. Anderson

Abstract

The University of Washington committee on instructional development and innovation is evaluated with respect to their effectiveness as a granting agency. The paper begins with a section defining evaluation and then applies that definition in a conceptual and empirical critique of funded activity. Conclusions drawn in the report include: (1) evaluation should not be required of one-time funding for new course development, (2) the committee should provide the grant receivers with specific evaluation requirements, (3) the committee should accept the responsibility for evaluating the overall funding pattern they establish, (4) the conceptual base of media-related evaluation needs to be changed, and (5) funding for the improvement of existing courses should extend through more than one grant.

Educational Assessment Center Project: 551

An Evaluation of the University of Washington Innovative Fund

Edwin R. Anderson

"...everything excellent is as difficult as it is rare."

Spinoza

The committee on instructional development and innovation at the University of Washington has been structures so as to allow the committee influence over instruction quality via judicious control of financial awards to teachers. The lack of monetary influence has often been stated as a serious handicap in the search for improved teaching; thus the innovative fund program provides an excellent opportunity to examine the merit of the control of awards concept. This report is concerned (1) with providing a critique of the evaluation activity associated with funded projects, (2) with reporting in summary form the results of some Educational Assessment Center (EAC) evaluation efforts, and (3) making recommendations to be considered by the fund committee in future, resource-allocation decisions. All of this indirectly assesses the usefulness of the control of awards idea. It is our sincere hope that the report will contribute to strengthening the impact of the fund and hence to improving the quality of student learning at the University.

The paper is divided into four major sections: (1) What is Evaluation?, (2) A Conceptual Critique of Fund Evaluation, (3) Empirical Results, and (4) Recommendations for Future Funding. The section discussing evaluation is included because, while most people are convinced that evaluation is necessary, and, in fact, engage in evaluation activity, not many people can explicitly describe the evaluation process. The literature on educational evaluation simply assumes that everyone understands the meaning of its central term, value, and gallops into discussion of how evaluation is to be done. The result of

neglecting the conceptual underpinnings of evaluation has been a situation in which everyone says we ought to have some evaluation (we agree, opposing evaluation as a concept is as bad as opposing motherhood), but no one is really sure what it is that is wanted. The section on evaluation seeks to clarify the nature of valuation and to use the concept of value adopted as a frame of reference for the other three sections of the paper. The definition used in this paper identifies valuation as a conceptual and logical activity which leads naturally into the second section of the paper, A Conceptual Critique of Fund Evaluation. While evaluation should not be confused with empirical method, it does draw heavily upon experimental techniques, hence the third section which reports the results of empirical work conducted by EAC. Because evaluations conducted by EAC of nonfunded courses have bearing on funded courses, the empirical section cites a range of EAC studies, funded and nonfunded. The final section contains a summary of the observations from the conceptual and empirical analysis in recommendation form.

What is Evaluation?

Philosophers of logic frequently rely on two terms, intension and extension, in describing the use of concepts. The extension of a concept is composed of a collection of items, a group of abstract ideas, or a set of events which fall within the field of applicability of the concept. Intensions are established by defining the concept in question with words or symbols. Different intensions may have the same extension. For example, in a particular room of twenty people, the intensions established with "the woman in the red dress" and 'Sally's mother' is the only woman wearing a red dress. In this context the two intensions share the same extension; in other contexts they would not. In logic, the

elaboration of extension leads to set theory and serves as the logical foundation of mathematics (Langer, 1953); the elaboration of intension leads to a theory of value (Hartman, 1967).

In order to use a term precisely, it is necessary to give the term a formal definition. Hartman (1967) has proposed the following definition of "goodness,"

> "...a thing X is a good C if and only if (1) X is a C, (2) C has the intension Φ consisting of predicates ϕ, (3) X has all the properties corresponding to ϕ." p. 154.

The definition is not nearly as ominous as it first appears. For example, in order for object X to be a good car, it must first of all be a car (1). The object will need to have four wheels, be self-propelled, etc. Secondly, there must be a set of properties which define the intension of C, the concept car (2). Φ is a symbol naming that set. We have then a list of predicates (ϕ) which serve as elements of the intensional set. Our complete concept of a car may state that a car gets "30 miles to the gallon of gas", "carries four passengers", "is inexpensive to repair", "costs less than $4000", etc. Each of these characteristics is part of the predicate list and is represented by the variable ϕ. Finally X will be a good C if it has each of the properties listed. Under this intension for car, a Honda is a good car while a Cadillac is not.

> "A thing is good insofar as it exemplifies its concept. This means that (1) the thing has a certain name, (2) this name has a meaning defined by a set of properties, and (3) the thing possesses all of the properties contained in the meaning of the name." (Hartman, 1967, p. 103)

The definition of a good C has a number of important implications.

> "...a difference had to be made between definitional properties and expositional properties. The definitional properties must in all cases be fulfilled, the value differentiations adhere to the expositional properties exclusively." (Hartman, 1967, p. 178)

If we enter a third object which we call a car (definitional properties fulfilled) into consideration, we will differentiate its value with respect to the other two by using the expositional properties (gas mileage, number of passengers, repair expenses, cost, etc.). This new car is a "fair" car if it has some of the expositional properties, is a "bad" car if it lacks some of the properties, and is a "no good" car if it lacks all the expositional properties. Value analysis is needed because naming is imprecise; the concept of car has too many objects in its field of applicability. Further differentiation is needed so we rank objects by including expositional properties with definitional properties. In addition to "good", "fair", "bad", and "no good", Hartman's approach allows us to define "better".

> "X is a better C than Y means that X has more expositional properties of C than Y and is therefore 'more of' a C than Y. 'Better than' in other words, relates two members of the same class, the first of which has more of the class properties than the second." (Hartman, 1967, p. 162)

The descriptive capability of empirical research is used to establish the properties of X and Y. The exact content of the list of expositional properties cannot be, however, empirically determined.

A second implication of Hartman's definition is that an object is both good and bad, valuable or not valuable, depending upon the concept being used. An object good under one concept may certainly be bad under another. Since concepts vary, value must not be taken as given in the object but rather must be seen as created by the person or persons holding the concept. Value is always value to someone.

The third implication is that "...value is rational. I can value a thing only if I know it, that is, if I know its name and its properties." (Hartman, 1967, p. 109). For this reason we call in an expert when we are having value problems. He knows more about the thing than we do, hence is more able to value it. Knowledge and valuation are thus intimately related. (Note 1: When we speak of people's motivations we often refer to their values as if they were central to motivation; this common rationale sees values as distinct from thinking. Hartman's definition removes this distinction. Note 2: Moral philosophy is now heavily influenced by the writing of John Rawls. He has equated goodness and rationality (See chapter VII of A Theory of Justice) without drawing upon Hartman's analysis.)

In the writing which follows this definitional section, there are two levels of evaluation, (1) evaluation of particular instructional settings and (2) evaluation of the particular evaluations. The instructors given innovative grants filed reports evaluating the results of their improvement efforts, those reports are in turn evaluated in this paper. Valuation is refexive in nature; it is reasonable to speak of evaluating an evaluation (there is no theoretical reason to stop at just two levels, evaluation is an infinite regress). A given concept with intension Φ_1, with its corresponding properties (ϕ), can be used to value an object selected by its definitional properties. We may in turn ask whether that concept is the appropriate concept for use in valuing the object; to do so we need another concept with intension Φ_2 to value the first concept The valuing concept at one level of analysis becomes fact for another concept at a second level of analysis. (See Anderson, 1975c, for further discussion of the structure of evaluation.)

The Greek "polis" was an organization of people who gathered in order to

establish mutual influence (Arendt, 1958). The polis member was expected to put forth his views on the course of action the city state should adopt and to support those views with reasoned argument. In many cases the members of the polis created reason by agreeing to a particular point of view. The political role derived from the polis is an action role; the politician enters an action arena where he attempts to influence the decisions of a group of similar politicians. The concept of evaluation thus far advanced is clearly related to political activity. In choosing a thought structure which can be applied to a real setting, i.e., the project being evaluated, the evaluator is entering into the action arena surrounding that setting. The trappings of science, such as statistical and research methodology, that the evaluator brings with him should not lead us to believe he is an objective, unbiased participant in the setting. Scientific approaches can sometimes establish which of two instructional methods leads to better performance on a test, but such approaches cannot determine whether that is the appropriate test or whether an alternative should be considered. Science can help us establish what is happening, but it can't help decide what should be happening. However, since most programs are designed to produce real world consequences, the evaluator stands with one foot in the political realm and one foot in the empirical realm. He is concerned with the structure of thought, i.e., the justification, associated with events he seeks to evaluate and he is concerned with empirical statements about those events. Valuation occurs when he brings the expositional properties of the thought structure together with his observations.

The foregoing discussion of evaluation frames the next two sections, A Conceptual Critique of Fund Activity and Empirical Results. The conceptual critique is our entry into the overall action arena of the fund; the empirical results report our entry into the workings of specific courses.

A Conceptual Critique of Fund Evaluation

Fund impact, from a strict development point-of-view, has been substantial. New courses are present on campus, new workbooks and texts are in evidence, and considerable media development has occurred. Most of the activity contracted for by the proposal writers has been carried out; while some of this work would have been done by highly motivated teachers without fund assistance, the existence of the fund has probably promoted interest in course development projects. The question that remains is, "What is the value of these activities promoted by the fund?"

There are two approaches to the valuation of development activity. The first of these, which we will call course improvement evaluation, involves the establishment of "better than" in the sense of Hartman's definition (p. 4). If, for example, a conversion from traditional instruction to self-paced instruction is planned and the content of the course is not being changed, then student achievement may be the measure of choice and evaluation is clearly addressing a "better than" question. Student achievement should be assessed both in the traditional course and in the changed course with a measure common to the two courses. It is the common content and the common measure which establishes the membership of both courses in the same conceptual class and hence makes them comparable under the common expositional properties that Hartman identifies as central to the "better than" type of evaluation. A second, independent and important measure is assessment of student attitude toward instruction via student ratings. This information can also be collected before and after change.

Most of the proposals which involved attempts to improve existing courses did not include provision for before change and after change assessment of achievement. The failure to collect data appropriate to evaluation of student

achievement and student attitude is a serious problem. The usual result in educational research, when variables thought to influence instruction are rigorously examined, is no significant effect (Dubin & Taveggia, 1968; Getzels & Jackson, 1963; Stephens, 1967; Wallen & Travers, 1963). The committee would be well advised to assume that development efforts funded by the committee will not result in real improvements and thus place the burden of proof for positive results squarely upon the grantee.

The policy statements of the committee include one asking for "clearly delineated plans for the evaluation of the project." This policy statement is particularly important since justification of fund activity depends upon the contents of the grantees' evaluations. Unless funded activities which propose to improve existing courses do indeed show course improvement, the fund itself cannot receive a favorable evaluation. The evaluation sections of most of the proposals are vague and usually do not focus attention on student performance as strongly as they should. The weakness of the proposal evaluation sections is not surprising as most of the grant writers are not used to doing educational evaluation. The committee has then the problem of educating its grant writers with respect to evaluation and should seek to remedy this problem by providing prospective grant writers with a pamphlet explaining the need for evaluation, suggesting procedures for evaluation, and finally, providing example grant evaluation sections. The potential grant writer should also receive copies of reports which meet the standards of evaluation expected by the committee. (Part of the efforts of EAC have been directed toward the development of techniques individual instructors can use to evaluate their own instructional changes. See Anderson, 1975 (b,c), plus reports soon to be issued. The educational research literature is weak where single classroom research techniques are concerned.)

Given the dismal record of efforts at educational improvement reported in the education literature, the fund committee must insist upon a strong evaluation program if they are to justify the continued existence of the fund in the instructional improvement aspect of development.

After having read the reports coming from the recipients of funds, we can but mourn opportunities lost. A number of the proposals could have been structured to permit, at the very least, quasi-experimental (Campbell & Stanley, 1963) tests of the changes implemented in the funded course. In the proposals and in the reports, the position was often taken that the merit of a change was beyond doubt whereas in truth the causes for doubt are legion and the empirical evidence is lean (See the empirical section of this paper).

The second approach to evaluation of development activity is less empirical and more conceptual; we will call it achievement pattern evaluation. The decision to add previously untaught courses to a department's offerings grows out of rational considerations within the academic concern of the department. The change in curricula represented by a new course is likely to be one of the most substantial impacts of the innovative fund. In a thoughtful article comparing various curricula, Walker & Schaffarzick (1974) conclude that "...different curricula are associated with different patterns of achievement." Whether a different pattern of achievement is better than the old pattern depends upon the conceptual systems one employs to evaluate the two patterns. The grant writer supplies a rationale for why the new course he is proposing should be funded. The committee decides which of several proposals of this type should receive funding. It is up to the committee members to provide a conceptual scheme which will guide the overall achievement pattern they choose to fund across the university. That conceptual scheme will establish the value of the achievement pattern resulting from funding

decisions and hopefully would encourage course development proposals consonant with it. The evaluation report associated with this type of development should thus come from the fund committee and not from the individual grantees. (The overall report filed by Kathleen Kaempki is descriptive and not evaluative.) In addition to the committees university wide perspective, the department offering the new course might be asked to provide a rationale justifying it, i.e., the department could explain how the content of the course fits the structure of knowledge in the field of concern. The grantees' evaluation function in this type of development is primarily an accounting role. He must verify that the course was developed and taught and that the money from the fund was spent as the proposal said it would be. The grantee is not in a position to establish the goodness of his single course from within the framework of the course itself. That type of evaluation must be reserved for within course improvement efforts as previously discussed.

The foregoing discussion leads to the conclusion that some types of proposals should not be subject to the policy statement requiring clearly delineated plans for the evaluation of the project. If an evaluation is desired from the grantee with a new course, it can be no more than a <u>critical</u> case history of the course; the grantee should be informed that this is all that is expected. Providing the grantee with high-quality (in the committee's judgment) example case histories would help. In practice, most grantees followed the case history approach when reporting back to the fund. Unfortunately the two dominant themes of those reports seem to be a repetition of the proposal and a reliance on the ratings of selected students. The student ratings are problematic in this case because of the strong possibility that the professors funded are more effective than normal teachers (perhaps associated with the high interest in teaching

shown by the grant writers) and because most of the new courses surely gathered students already biased in favor of the subject matter since they were predominantly voluntary enrollment courses. While comparison of innovative courses with overall university ratings is of interest, this should be viewed as weak evidence in favor of the funded changes.

In the project entitled, Planning the Evaluation of Instructional Innovations, EAC proposed to accumulate data on how effectively EAC staff could work with innovative faculty. Our research data is found in the evaluation reports issued by the center (See Gillmore & Sprenkle, 1975; Sprenkle & Gillmore, 1975; and Anderson, 1975b, for examples) but we would like to add a subjective impression. We had a concern at the beginning of our evaluation efforts with the degree of receptiveness of the faculty to evaluation; we expected to encounter resistance, i.e., that psychological euphemism for "he refuses to play the game by my rules." In fact, the faculty members who have asked for our support have been very open to our advice and have in general been quite satisfied with our assistance. However, in spite of encouragement from the fund committee, not many faculty who were seeking funding were among those who consulted with us. In keeping with the distinction made earlier between grants seeking to improve existing courses and grants presenting new courses, the recommendation is made that grantees seeking to improve the achievement of students in existing courses should be required to present more rigorous plans for evaluation. In order to increase the standards of the judgmental process for determining the adequacy of an evaluation section, we recommend the addition of a faculty member trained in evaluation to the fund committee. Grant seekers with proposals for the improvement of existing courses should be required to consult with the committee evaluator or with the staff at EAC. In view of our experience thus far, the require-

ment for evaluation assistance should not prove unduly onerous to the faculty (there are, however, sample selection effects in our experience since all faculty dealt with to this point have been volunteers). We see no need for such consultation for proposals establishing new courses.

Empirical Results

Media Development. There were numerous proposals which asked for money for course-related media development, but there was very little data collected on the effectiveness of the media once it was developed. There can be no doubt that students can learn from media (e.g., slides, tapes, films, etc.) but there is little evidence in the educational research literature to support the enthusiasm of many faculty for media over other means of instruction.

At least one grantee, Professor Lauritzen (Lauritzen & Daniels, 1975), has reported on the effectiveness of adding media (in the form of TV tapes) to his course (Engineering). In his initial proposal, Professor Lauritzen suggested that the TV tapes would be motivational. Two types of data were analyzed to check that hypothesis: (1) students were asked for estimates of the amount of time they had studied for a unit of the course before being given the unit examination and (2) the performance of the students on the examination. Comparison of TV watchers with nonwatchers on both measures failed to support the hypothesis even though both are plausible indicators of increased motivational levels.

A second professor (Thomas in Economics 200) has reported the use of a standardized economics test for assessing student achievement. His students scored above the mean on this test and he attributes a portion of their success to his extensive use of media. He also reports that the students gave high ratings to the media and the course. We took the liberty of sitting in Professor Thomas' class for three of his lectures. Our observation convinced us that the

method he used for assessing media effectiveness is invalid. We have no quarrel with the test he used even though we have not seen it; the test is most likely an adequate measure. The problem is with the failure to rule out probable confounding variables. Evidence from the educational research literature has shown clarity of instruction to be positively related to student achievement (See Rosenshine & Furst, 1971, for an appropriate review). Dr. Thomas is an excellent lecturer; his verbal presentation has all of the characteristics of lectures leading to higher student performance. In order for him to claim the high achievement of his students as media related, Dr. Thomas would have to gather baseline data from his course when the media were not used for comparison with his course when the media were present.

At an Innovative Fund-EAC Instructional Improvement Seminar, Dr. Thomas stated that a major reason for using media was the massive inattention that he had observed in the large lecture classes in Kane Hall and offered the corresponding hypothesis that media inclusion would reduce the problem. We agree that inattention is a readily observable problem in Kane and that attention is a variable closely related to learning (Loftus, 1972; Morsh, 1956; Bloom, 1974). Unfortunately Dr. Thomas did not systematically observe attentional behavior before and after the development of the media for Economics 200. We might also ask, "What is the student paying attention to?" One movie was prepared to show that joint ownership can lead to conflicts. The movie developed the theme by showing two "dudes" happily sharing a jointly purchased pizza until a conflict over the last piece ended their beautiful relationship. We doubt that the point required the expense of a movie to demonstrate it. After a statement that kittens are cute, color slides of kittens were shown to illustrate the point. Several color slides were also shown after a similar statement about babies. From an

entertainment perspective the slides were well done, but the attention of the students was not necessarily drawn to the point being made from an economic perspective nor is it likely that the point was made in the most economical manner consonant with student achievement. Achievement is not the only measure of importance in educational settings, but it deserves more attention than it has been getting. The students' qualitative ratings about how the learning situation has affected them (student ratings data) are also important but we need to carefully examine the argument that high ratings and media presence go together. We believe that Dr. Thomas would deserve and would get high qualitative ratings with no more media present than a blackboard or overhead projector.

A number of reports received from grantees with media related funding included information about efforts to assess the quality of the developed media. Recall, however, that Hartman's definition of value assessment requires that we know the conceptual scheme used in valuation. Since none of the reports made reference to achievement measures, the conceptual scheme underlying quality judgments would appear to involve such concepts as clarity of images, camera techniques, color vs. black and white, etc. In short the judgments are related to concepts of media professionalism much like those the committee for selecting the academy awards might use. In spite of frequent assertions made by faculty that they are competing with television, we do not believe that the standards applied to television are appropriate for instructional media. Television seeks to entertain whereas instructional media seeks to teach. Although entertaining and teaching are not necessarily mutually exclusive, media developed for the two purposes tend to have cost differences. Entertaining usually means color production and other expensive embellishment. Evidence collected in learning

studies has shown that an inexpensive black and white film with several line drawings is as effective as the same film with more elaborate drawings and color photography (May & Lumsdaine, 1958). Recognition memory for pictures is as good for line drawings as it is for color photos of the same scene (Nelson, et al., 1974). A major justification for media development is that media can be used to free the instructors time and that several instructors can use the same media. This is basically a cost related argument and the strength of the argument is considerably weakened by including frills in media which are more related to entertainment than to instruction. Secondly, a cost-benefit analysis requires assessment of benefit, in this case assessment of student achievement. Until this achievement data is presented, the true cost of media cannot be determined.

The addition of media to a course can be researched in a manner which will allow evidence to be collected to address the questions we have raised concerning media cost and media effectiveness. Any future grants for media development should include a detailed statement of the experimental procedure that will be used to provide evidence bearing on the media-related issues raised in this section.

Semi-Self-Paced Instruction. The personalized system of instruction (PSI) proposed by Keller (1968) has the following features:

1. Students are allowed to proceed through the course at their own pace.
2. Each student must reach an 85% or 90% level of mastery on a unit test before being allowed to move to the next unit. Repeated testing occurs until mastery is attained.
3. Students are allowed to attend lectures and demonstrations

as a reward for finishing the appropriate unit. Lectures are not seen as containing critical information.

4. Learning occurs primarily from the written word.

5. Proctors are used to test and consult with students.

While the overall PSI plan has frequently been shown to be more effective in promoting student achievement than traditional lectures (See Sherman, 1974, for collected papers), there has been little research to determine which of the features is critical to the success of the method. Instructors on this campus have modified the features while, perhaps falsely, assuming they were still getting the improvement benefits advertised for PSI.

Engineering 141, introductory FORTRAN programming, has been taught by several professors in a modified PSI format. The students were given frequent tests, they were allowed one repeat on the test if they did not get 90% correct on the first test, and learning was primarily through the written word. Two classes taught by this method were given the same final examination as five traditional classes (i.e., lectures, infrequent testing). There was no significant difference between the two types of classes (Anderson, 1975b). Professor Dunn of one of the semi-PSI classes was given a small grant to assist him in finishing a workbook containing three types of FORTRAN practice problems. The sample multiple-choice problems were similar to (and for 50% of the problems were identical to) the multiple-choice problems used in unit examinations. Students in a fall section of his course received the workbook while students in a winter section did not. There was no difference in the test performance of the two groups even though the students rated the workbook favorably (Anderson, 1975b).

By comparing the proportion of students getting an item correct across the

many classes tested, we discovered that problems easy for one class were likely to be easy for other classes and questions hard for one class were likely to be hard for the others (correlations on the order of .70 to .80 were found). The classes have common teaching difficulties. We tried to capitalize on this by giving students a written statement of what they were to learn, a brief written explanation for the statement, and a practice problem for the statement; this was only done for one-half of the difficult test items (determined from a prior quarter). The students were told that the handouts they were receiving were directly related to difficult questions on the unit examination; they were asked to pay special attention to the handouts. The result was again no improvement in achievement although the students gave high ratings to the presence of the handouts.

Tom Love received a grant from the innovative fund for the purpose of teaching structured programming concepts in the introductory FORTRAN class. His students did not perform any better on the final examination than the traditional and semi-PSI classes. (His class did have less time for the final examination than other classes.) He did, however, collect data which provide a possible clue to the lack of success in these improvement efforts. He kept track of the number of times students attempted to run their programs before they were able to run the program successfully. From this data, Love concluded that the students were requiring large amounts of time in order to solve simple programming problems (some students took up to 30 or 40 trials). It appears that the main behavior of the students in the FORTRAN course is writing and debugging programs (students say the course is quite time consuming). Efforts at changing their learning through workbooks and testing have possibly been peripheral to what the student is actually doing. The programming student is not, as is usual, studying for tests.

Programming is an ideal place to study learning since the students leave observable tracks in their learning process. The error runs can be conveniently collected for analysis. With the data thus far collected and further additional effort focussing on student programming efforts, it might be possible to produce an improvement in FORTRAN instruction.

Experience with the concentrated effort to improve the FORTRAN course in combination with other experience leads to two observations, (1) hard work by itself does not lead to improvement and (2) repeated improvement effort may be required before an improvement project can show positive results. Even extended effort may not prove successful. These observations lead to a recommendation to the funding committee: the committee should be prepared to fund fewer course improvement projects over a longer period of time because a serious improvement effort is likely to require more than a one-time infusion of funds into a course. This recommendation does not apply to one-time grants for new courses or achievement pattern development.

Many instructors have reported student ratings as part of the evaluation sent back to the innovative fund. While student ratings data is valuable, the way in which it is collected leaves much to be desired. Ratings from a single course are not an adequate source of data since they offer nothing for comparison. Semi-PSI courses tend to receive good ratings but Sprenkle & Gillmore (1975) showed that care must be exercised in interpreting those ratings. Seven semi-PSI courses at UW were rated more highly than general university courses using the same evaluation form, however, when ratings of those seven courses were compared with ratings from seven traditional sections of the same courses, no differences were found on items common to both sets of evaluations. These data, combined with the high ratings given to the FORTRAN workbook and handouts

even though achievement improvement could not be attributed to the workbook and handouts, clearly point to the inadequacy of the major form of data (student ratings) reported back to the funding committee.

Evaluation of Courses Designed to Influence Student Attitude. EAC has been involved with two courses that sought to influence students' attitudes. In one, students were given a scale measuring attitudes toward disabled persons (ATDP) at the beginning and at the end of the course. There was no difference between the pre and post ATDP scores; analysis of the pre scores offers a reason for the lack of difference. The mean of the pre-course ATDP scores was 125.37. This figure should be compared with normative data reported in Yuker, Block & Younng (1970): the mean ATDP score for males is 106.65 and for females is 114.18. A normative sample of disabled persons only had an average score of 122 on the ATDP scale. The failure to produce change in attitudes is due to the presence of an initial attitude ceiling effect. In the second course, an attempt was made to favorably (concept not well defined by instructor) influence student attitudes toward suicide (Gillmore & Sprinkle, 1975). Attitudes toward suicide were measured with a Semantic Differential technique whereby students responded to the word suicide with ratings on twenty-one, seven point bi-polar scales, e.g., sad-happy, aggressive-meek, brave-cowardly. The students showed no change in their ratings from pre to post measures. There was no normative data for the suicide scale as it was prepared solely for this class. It is worth noting that 67 of the 71 students answered yes to the question, "When enrolling, was this a course you wanted to take?" The lack of an attitude change in either course may be due to strong subject selection effects associated with volunteering for elective courses. It is also noteworthy that neither instructor oriented their attitude change effort with a theory of

attitude change. Evaluation sections of proposals seeking funding for attitude change efforts should contain (1) provisions for pre-course and post-course assessment of the attitude, (2) provision for obtaining normative data from students who didn't volunteer for the course, and (3) provision for exploring the connection between the change effort and a theory of attitude change.

Summary of the Recommendations for Future Funding

The following numbered statements are a summarization of recommendations made thus far. The committee should...

1. ...categorize proposals into two types, (a) course improvement proposals and (b) achievement pattern proposals. The committee should ask for rigorous evaluation of course improvement efforts and collect case histories for the evaluation of achievement pattern proposals.

2. ...consider providing grant seekers with evaluation information appropriate to the type of grant they are seeking. The committee must educate the grant seekers with respect to evaluation if the fund is to have a strong evaluation program associated with it.

3. ...accept the responsibility for evaluating the overall pattern of achievement they have funded.

4. ...add an evaluation specialist to the committee and ask grant seekers to consult with the evaluator.

5. ...request more rigorous evaluation efforts from faculty seeking media-related grants. Student achievement should be related to media development. Additionally media costs should be lowered by removing entertainment-oriented professional standards from instructional media.

6. ...adopt the belief that course improvement is difficult to produce and document. In accordance with this belief, course improvement grants should cover

longer time periods than achievement pattern grants.

7. ...encourage abandonment of student ratings as the major type of reported data. Grantees should be encouraged to use achievement data primarily and student ratings as secondary data.

8. ...request before and after measures in courses seeking to produce student attitude change. Normative data from populations randomly selected should also be requested to control for the bias introduced by students selecting the funded courses as electives.

Evaluation has been too long in coming to the individual instructor's classroom. If the fund committee encourages grant seekers to take self-evaluation seriously and if models and methods can be given to the grantee, then maybe evidence can be generated which will lead to and document real change for the better. However, the "better" in the last sentence depends upon the rigor and general acceptability of the conceptual analysis which leads to it. The strength of the funds real impact on the learning of students lies in the goodness of the conceptual base the fund committee lays for evaluation; without such a base funded activity will be all effort and no improvement.

References

Anderson, Edwin R. Evaluation: The justification of practice. University of Washington Educational Assessment Center technical report, project 297. May, 1975, 21 pages. (a)

Anderson, Edwin R. The certainty of information in instructional decision making. University of Washington Educational Assessment Center technical report #76-4, project 503. August, 1975, 23 pages. (b)

Anderson, Edwin R. Personal inquiry in the classroom: An alternative approach to educational research. University of Washington Educational Assessment Center technical report #76-5, project 4. August, 1975, 9 pages. (c)

Arendt, H. The Human Condition. Chicago: University of Chicago Press, 1958.

Bloom, B.S. Time and learning. American Psychologist, 1974, 9, 682-688.

Campbell, D.T. & Stanley, J.C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed), Handbook of Research on Teaching. Chicago: Rand McNally, 1963, 171-246.

Dubin, R. & Taveggia, T.C. The Teaching-learning Paradox: A Comparative Analysis of College Teaching Methods. Eugene, Oregon: Center for the Advanced Study of Educational Administration, University of Oregon, 1968.

Getzels, J.W. & Jackson, P.W. The teacher's personality and characteristics. In N. L. Gage (Ed), Handbook of Research on Teaching. Chicago: Rand McNally, 1963, 506-582.

Gillmore, G. M. & Sprenkle, V. E. An evaluation of T. F. Lewin's Social Work 470. University of Washington Educational Assessment Center technical report #76-6, project 504. September, 1975, 10 pages.

Keller, F. S. "Goodbye, Teacher..." Journal of Applied Behavior Analysis, 1968, 1, 79-89.

Hartman, R.S. The Structure of Value. Carbondale: Southern Illinois University Press, 1967.

Langer, Suzanne. An Introduction to Symbolic Logic. New York: Dover Publications, 1953.

Lauritzen, P.O. & Daniels, P.D. Video tape lecture demonstrations in a PSI course. Paper filed as a report to the innovative fund, 1975.

Loftus, G.R. Eye fixations and recognition memory for pictures. Cognitive Psychology, 1972, 3, 525-551.

May, M.A. & Lumsdaine, A. A. Learning from Films. New Haven: Yale University Press, 1958.

Morsh, J.E. Systematic Observation of Instructor Behavior. Development report, AFPTRC-TN-56-52. San Antonio, Texas: Air Force Personnel and Training Research Center, Lackland Air Force Base, 1956. Cited in R. C. Anderson. Control of student mediating processes during verbal learning and instruction. Review of Educational Research, 1970, 40, 349-369.

Nelson, T.O., Metzler, J., Reed, D.A., & Gorchaff, D. R. Role of details in the long-term recognition of pictures and verbal descriptions. Journal of Experimental Psychology, 1974, 102, 184-186.

Rawls, John. A Theory of Justice. Cambridge: Belknap Press of Harvard University Press, 1971.

Rosenshine, B. & Furst, N. Research on teacher performance criteria. In B. O. Smith (Ed). Research in Teacher Education: A Symposium. Englewood Cliffs, N.J.: Prentice-Hall, 1971, 37-72.

Sherman, J.G. PSI Personalized System of Instruction: 41 Terminal Papers. W. A. Benjamin, 1974.

Sprenkle, V.E. & Gillmore, G.M. An evaluation of mastery instruction courses by use of student ratings. University of Washington Educational Assessment Center technical report #76-3, proje t 503. July, 1975, 30 pages.

Stephens, J. M. The Process of Schooling. New York: Holt, Rinehart & Winston, 1967.

Walker, D.F. & Schaffarzick, J. Comparing curricula. Review of Educational Research, 44, 1974, 83-111.

Wallen, N. & Travers, R.M.W. Analysis and investigation of teaching methods. In N. L. Gage (Ed), Handbook of Research on Teaching. Chicago: Rand McNally, 1963, 448-505.

Yuker, H.E., Block, J.R. & Younng, J.H. The Measurement of Attitudes Toward Disabled Persons, RD-834, INA. Mend Institute at Human Resources Center, Albertson: New York, 1970.